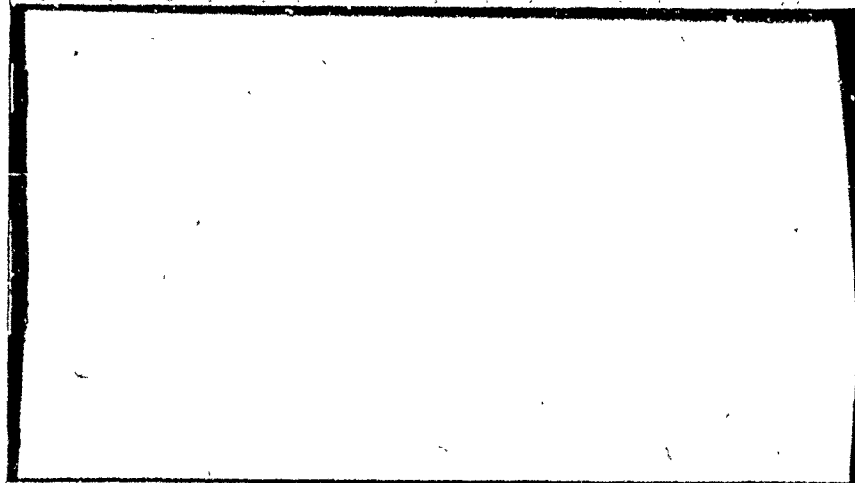


(P)

AD A 1 1 4 1 2 2



DTIC FILE COPY

DTIC
ELECTE
S MAY 5 1982 D

Prepared by
Bell-Northern Research Ltd.
Ottawa, Ontario

A

This document has been approved
for public release and sale; its
distribution is unlimited.

82 05 04 082

APRIL 1982

FINAL REPORT

Audio Conferencing Using Dynamic Channel-
Assignment System

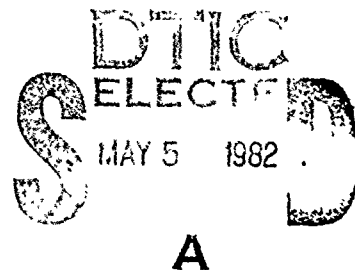
PREPARED FOR:

Defense Advanced Research Projects Agency

Contract No.: MDA903-81-C-0180
BNR Ref. No.: P379/TR6735

PREPARED BY:

Bell-Northern Research Ltd.
P.O. Box 3511, Station C
Ottawa, Ontario
K1Y 4H7




This document has been approved
for public release and sale; its
distribution is unlimited.

SUMMARY

This study examines dynamic channel assignment techniques to demonstrate their feasibility in audio teleconferencing. Computer simulation of dynamic assignment is used to investigate the rules for controlling the bit-rate and is used to subjectively evaluate the speech quality of audio conferencing system.

Efficient encoding of speech signals requires the removal of redundancies in the speech waveform. Predictive codings techniques are applicable to dynamic channel assignment. Adaptive Predictive Coding at 16 kbps is suitable for a high quality / high rate channel. The closely related technique of Linear Predictive Coding is appropriate for a low quality / low rate channel. A dual-mode coder combines both the above techniques together with a switching control algorithm. The control algorithm selects the mode of the dual-mode coder based on the relative amplitudes of the conference participants.

Computer simulations of a dual-mode coder indicate that the technique is well suited to audio conferencing. Dynamic assignment of the coding mode introduces only minor quality degradations when proper control algorithms are used. The overall impression is one of high quality coding for most of the duration of the conference.



Predictive coding methods are good candidates for this application. The similarity of coding techniques allows for a substantial sharing of hardware in a dual-mode coder implementation. Also, from a quality point of view, the mode transitions are graceful.

The main conclusion of this study is that dynamic channel assignment based on voice activity is technically feasible for use in the NCATS environment. A dual-mode coder based on predictive coding methods has shown itself to be particularly attractive for this application.



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
<i>Fuller on file</i>	
By	
Distribution	
Availability Codes	
Dist	Spec 1
<i>A</i>	

TABLE OF CONTENTS

SUMMARY	i
TABLE OF CONTENTS	iii
LIST OF FIGURES	v
1. OBJECTIVES	1
2. PREDICTIVE CODING	1
2.1 Background	1
2.2 Adaptive Predictive Coding (APC)	2
2.3 Linear Predictive Coding (LPC)	3
3. DUAL-MODE CODER	4
3.1 Configuration	4
3.2 Coder Structure	4
3.3 Switching Control Algorithm	10
4. IMPLEMENTATION - EXPERIMENTAL SUPPORT	12
4.1 BNR Simulation Facility	12
4.2 Hardware Support	12
4.3 Software Support	13
4.4 Simulation System Configuration	14

TABLE OF CONTENTS

5. RESULTS AND CONCLUSIONS	14
6. REFERENCES	18

TABLE OF FIGURES

Figure 3.1: Network Architecture for the Dual-Mode Speech Coder

Figure 3.2: Dual-Mode (APC/LPC) Speech Coder

Figure 3.3: Dual-Mode Coder Control Algorithm

Figure 4.1: Simulation of Dual-Mode Coder

1. OBJECTIVES

This study examines dynamic channel assignment techniques to demonstrate their feasibility in audio teleconferencing. Computer simulation of dynamic assignment is used to investigate the logical rules for controlling the bit-rate and is used to subjectively evaluate the speech quality of the audio conferencing system.

The objectives of the study are twofold:

- a. To study the effects of dynamically assigning increased channel capacity to the active speaker and reduced capacity to other participants by real-time bi-modal speech coding.
- b. To explore decision rules for channel-assignment based on monitored speech activity of three conference participants and network imposed delays on receipt of such channel-assignment information.

2. PREDICTIVE CODING

2.1 Background

Speech signals exhibit considerable redundancy when examined over short durations. Efficient encoding of speech requires the removal of these redundancies before transmission. Linear predictive coding [1] is a technique whereby the predictable components of the speech waveform are removed from the signal by estimating the current sample value as linear combination of previous sample values. This technique is effective because the characteristics of the speech signal are relatively constant for short segments (20-30 ms).

In predictive coding schemes it is often necessary to modify the coefficients of the linear combination in order to track the changing speech sounds. For speech signals the prediction coefficients are calculated 30 to 50 times per second in such a manner as to minimize the mean square error between the signal and its estimate. Typically 8 to 12 prediction coefficients are adequate. These coefficients are transmitted to the decoder for proper signal reconstruction.

The prediction error signal is known as the prediction residual. Predictive coding techniques differ primarily in how the properties of the residual are modelled and encoded for transmission. Adaptive Predictor Coding (APC) [2] and Linear Predictive Coding [3] are two closely related techniques which are suitable for use in a dual-mode coder.

2.2 Adaptive Predictive Coding (APC)

APC is the name for a class of coders in which the entire residual signal is transmitted from the encoder to the decoder. With no quantization of parameters, it is possible to exactly reconstruct the speech signal from the residual and the prediction coefficients. The goal of APC is to minimize the distortions incurred when the residual and the prediction coefficients must be quantized for transmission.

The transmission of the quantized residual requires approximately 14 kb/s of bandwidth. This relatively large bandwidth is necessary to transmit an accurate representation of the residual. The accurate reproduction of the residual at the decoder is the key to natural sounding speech. It is the details of the residual signal which contain

the information for generating high quality speech output.

APC at 16 kb/s has been selected for the high quality mode of the dual-mode coder. Subjective quality evaluations indicate that APC yields speech quality comparable to 6 bit log PCM. This quality is slightly less than the quality of conventional telephone channels (approximately 7 bits log PCM).

APC at 9.6 kb/s is possible at the expense of significantly more complex processing. In this version of the technique the residual signal bandwidth is reduced to about 8 kb/s by a combination of signal processing steps and vector encoding of the quantized residual. Current work on APC at 9.6 kb/s indicates that this reduction in bandwidth can be achieved with only minor reductions in speech quality. In this report we restrict our investigation to APC at 16 kb/s.

2.3 Linear Predictive Coding (LPC)

LPC is an alternative predictive coding scheme which models the prediction residual and transmits the parameters of the model. For voiced speech, such as vowels, the residual is modelled as a periodic impulse train. The spacing of the impulses is determined by the fundamental frequency of vibration of the vocal cords. For unvoiced sounds, such as fricatives, the residual is modelled by a random signal such as a Gaussian noise source.

LPC at 2.4 kb/s has been chosen as the low quality mode of the dual-mode coder. The LPC technique imposes certain quality limitations on the reconstructed speech waveform. These limitations derive from the

restrictions of the residual model. The binary classification of speech frames as either voiced or unvoiced is inadequate especially when classifying a voicing transition. Another limitation of the model is the classification of the excitation signal as being either an impulse train or a noise source. These classifications are not appropriate for all types of speech sounds. A third source of degradation is the use of a fixed pulse shape for the periodic excitation signal which ignores speaker dependencies. The result of these degradations is that the decoded speech sounds unnatural and exhibits buzziness or throaty characteristics.

3. DUAL-MODE CODER

3.1 Configuration

Figure 3.1 shows a block diagram of the audio conferencing network. Each site requires a single dual-mode encoder and one dual-mode decoder for each of the other conference sites. A network control unit controls the mode of the encoder and each decoder based on the activity of each participant.

3.2 Coder Structure

The dual-mode coder is illustrated in block diagram form in Figure 3.2. The linear prediction analysis block processes a frame of speech (30 msec sampled at 6.5 kHz) and produces a set of eight reflection coefficients and a frame of the prediction residual. The coefficients

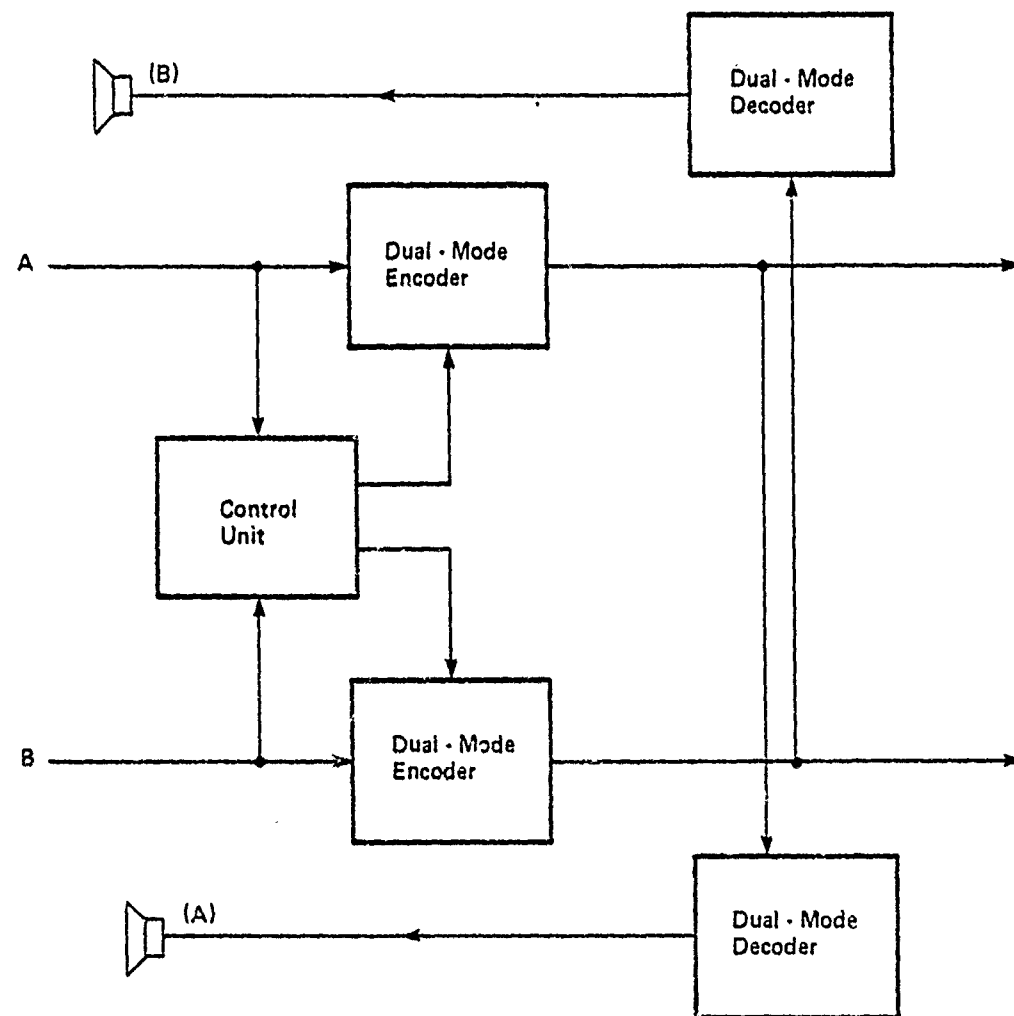


Figure 3.1 Network Architecture for the Dual-Mode Speech Coder

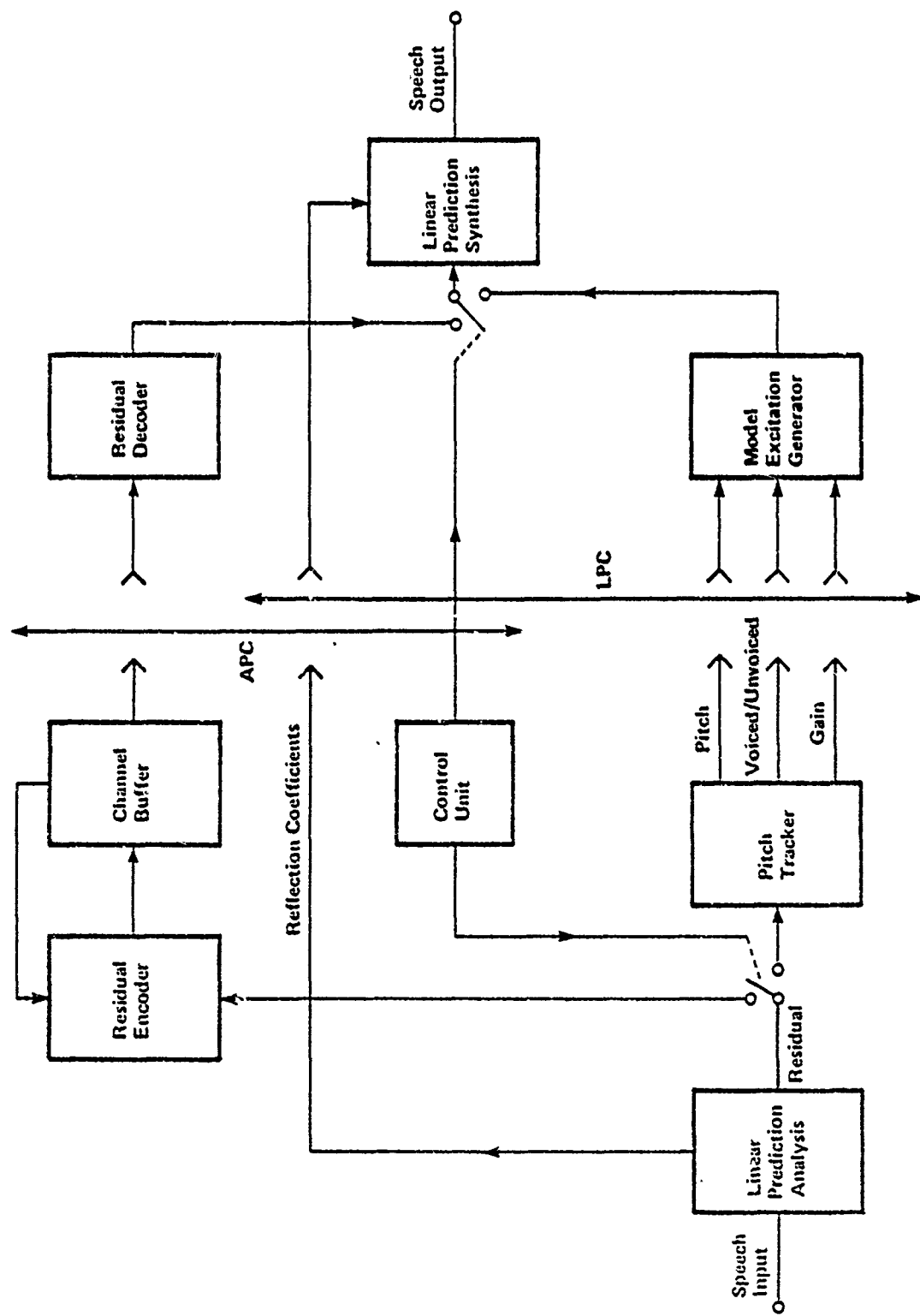


Figure 3.2 Dual - Mode (APC/LPC) Speech Coder

are obtained from the linear prediction coefficients by a one-to-one transformation. The reflection coefficients, however, are bounded and thus are more amenable to quantization. Both modes transmit the reflection coefficients to the decoder to be used by the linear prediction synthesizer for reconstructing the output waveform. The reflection coefficients are linearly quantized with the quantization parameters shown in Table 3.1. Approximately 1900 bits/sec are used to transmit these coefficients.

In the LPC mode the residual is used by a pitch tracker to determine several parameters of the input frame. The algorithm used is based on the "simplified inverse filter tracking" (SIFT) technique of Markel [4]. The pitch tracker decides whether the input is voiced or unvoiced. For voiced sounds an estimate of the pitch period is calculated. The RMS value of the residual is also calculated. The parameters are then quantized and transmitted to the decoder. Pitch and voicing information are allocated 8 bits and the RMS energy is logarithmically quantized to 6 bits.

The synthesizer uses the energy and pitch information to model the true residual. The approximated residual and the reflection coefficients are used by the synthesizer to construct the final speech output signal.

In the APC mode the residual itself is transmitted to the decoder. Entropy coding is used to code the residual. Entropy coding results in a variable bit rate encoding of the residual. The instantaneous number of bits per samples can vary from 1 bit per sample to more than 4 bits per sample. A channel buffer is employed to convert the variable bit rate to a fixed rate of less than 2 bits per sample. Channel buffer

overflow is prevented by controlling the rate of the residual decoder as a function of the buffer occupancy. At the decoder, the residual is reconstructed and used by the synthesizer to produce the speech output.

TABLE 3.1 Reflection Coefficient Quantization Parameters

Reflection Coefficient	Bit Allocation	Quantizer Range	
		Minimum	Maximum
K1	8	-0.9999	+0.9146
K2	8	-0.9287	+0.9937
K3	8	-0.9287	+0.8812
K4	8	-0.7264	+0.9751
K5	7	-0.7264	+0.8812
K6	7	-0.4204	+0.9609
K7	6	-0.7264	+0.7260
K8	6	-0.5985	+0.5985
TOTAL	58		

3.3 Switching Control Algorithm

The switching control algorithm selects the mode of the dual-mode coder as a function of the relative amplitudes of the conference participants. An important function of the control algorithm is its ability to prevent rapid switching back and forth between the two modes. To this end, the algorithm monitors both the current and past participant activity to select the proper mode.

The algorithm, based on [5], is a counting procedure which increments or decrements a counter for each comparison of participant amplitudes. If the selected speaker's level exceeds the highest level of the other participants by 3dB, his counter is incremented; otherwise it is decremented. If the counter reaches a threshold appropriate for an active speaker, the coder is switched into the APC mode. Incrementing beyond this count is suppressed, but decrementing may proceed until the count drops to an appropriate level to consider the speaker inactive. The coder is then switched into the LPC mode and remains in that state until the APC mode conditions are again satisfied.

The counter increment and decrement can be independently specified. Figure 3.3 shows the functioning of the control algorithm for a hypothetical set of inputs.

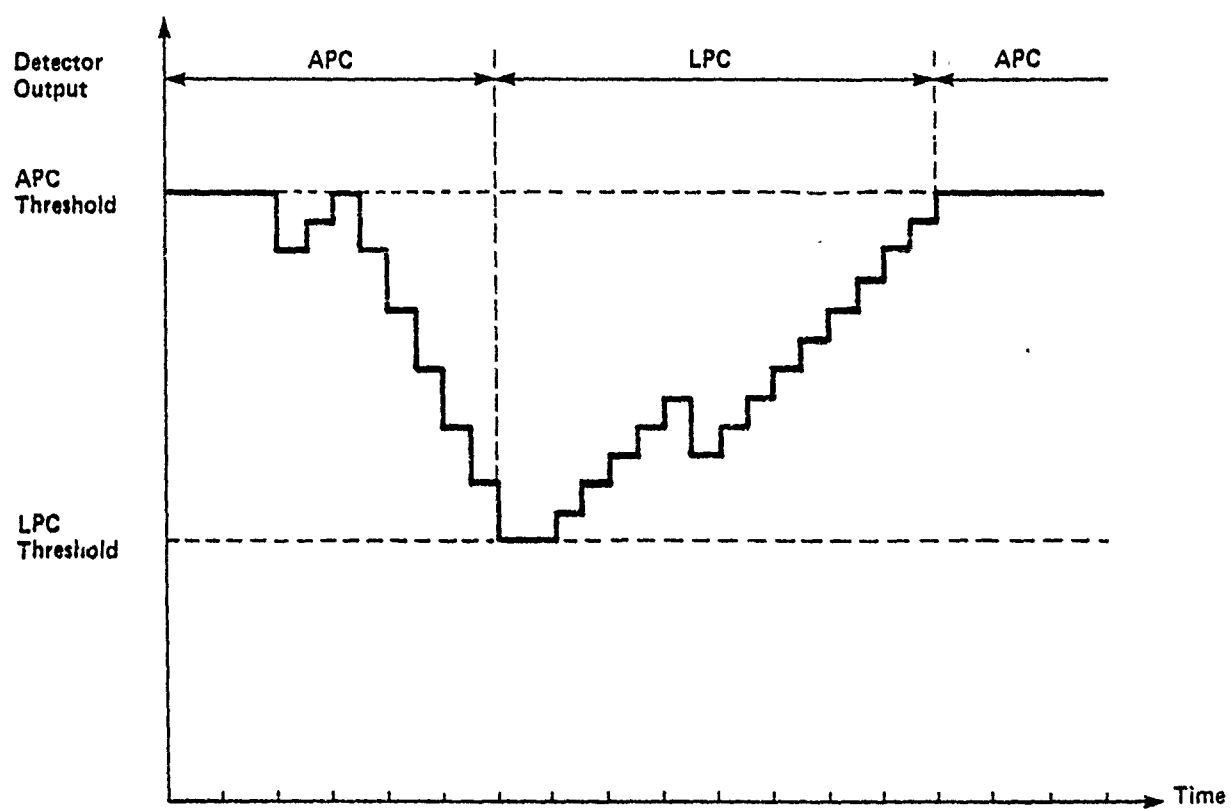


Figure 3.3 Dual - Moder Coder Control Algorithm

4. IMPLEMENTATIONS - EXPERIMENTAL SUPPORT

4.1 BNR Simulation Facility

The speech simulation facility includes two processors and several peripheral devices. A Digital Equipment Corporation PDP-11/45 supplies general purpose computational power and a Floating Point Systems' AP-120B array processor provides the high speed processing necessary for real-time simulation. The processors are interconnected via a standard PDP-11 UNIBUS and can share processing requirements. Either processor can act as a master with the other processor acting as a slave.

The facility is supported by standard computer peripherals as well as a speech interface. The speech interface contains 15 bit A/D and D/A converters. The sampling frequency of the interface is software selectable from 0 - 50 kHz. The digital hardware is augmented by analogue equipment such as filters, amplifiers and tape recorders.

4.2 Hardware Support

In order to simulate the audio conferencing system it was necessary to construct a device which could simultaneously monitor the speech activity of the conference participants. A signal level detector was designed and constructed to provide this function.

The signal level detector is a circuit that supplies the PDP-11 with amplitude information from a maximum of four input lines. The information can be a direct PCM encoding of one or several inputs or can be a PCM encoding of the output of a peak level detector associated with

each input. Sample values are represented by 8 bit numbers which can then be sent to the PDP-11 by a direct memory transfer. The sampling frequency is software programmable from 0 to 10 kHz. The number of input lines and the peak detector mode can be selected via control registers.

4.3 Software Support

The primary software effort was devoted to the real-time coding software. Additional software was necessary to support the signal level detector functions and to provide control of the dual-mode coder.

The real-time coder simulation required Adaptive Prediction Coding and Linear Predictive Coding to be implemented on the AP-120B array processor. Most of the software effort was expended in the conversion of the coding algorithms from FORTRAN simulations into array processor assembly language. Each algorithm was converted separately and tested on the array processor. The resulting array processor programs were then merged into a single dual-mode coder.

To complete the simulation system a device driver was written to provide the software interface between the PDP-11's operating system (RSX-11D) and the signal level detector hardware. A FORTRAN control program was implemented to control the real-time dual-mode coder on the basis of the amplitude information obtained from the level detector.

4.4 Simulation System Configuration

Figure 4.1 shows a block diagram of the system configuration. In the simulation only one participant's voice is coded by the dual-mode coder. The other participants serve only to provide information to the control unit. Each participant's speech activity level is measured by the signal level detector. The control unit monitors the activity levels to select the mode of the dual-mode coder.

5. RESULTS AND CONCLUSIONS

Computer simulations of a dual-mode coder indicate that the technique is well suited to audio conferencing. Dynamic assignment of the coding mode introduces only minor quality degradations when appropriate control algorithms are utilized. For example, rapid upgrading of the quality assigned to a speaker (less than 1 second) allows listeners not to be very concerned about the reduced initial quality. Thus, the overall impression is one of high quality coding for most of the duration of the conference.

The choice of the coding techniques for this application is significant. Adaptive Predictive Coding and Linear Predictive Coding are good candidates for two reasons. First, the similarity of the coding techniques allows for a substantial sharing of hardware in a dual-mode coder implementation. Second, from a quality point of view, the two techniques allow the quality transitions to be graceful. In fact, when simulating the dual-mode coder it was difficult to perceive the exact instant of mode transition.

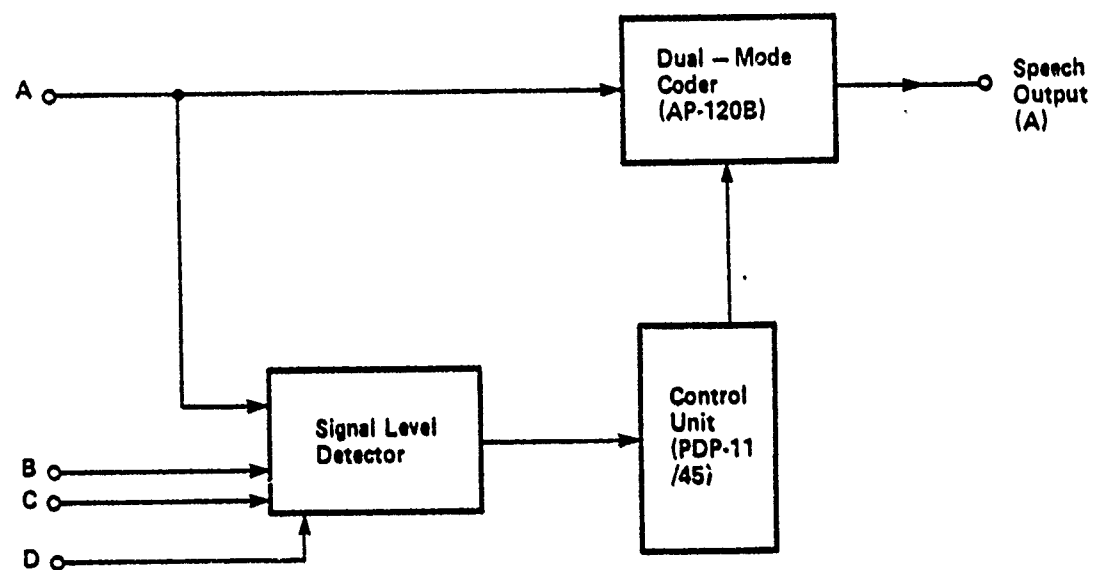


Figure 4.1 Simulation of Dual - Mode Coder

The advent of digital signal processing integrated circuits has made an economic hardware implementation of the dual-mode coder quite feasible. The Nippon Electric Company 7720 IC has been investigated for Linear Predictive Coding. We speculate that three of four of such chips performing the signal processing operations coupled with a microprocessor controller would be adequate for a hardware implementation of the dual-mode coder.

As stated previously, the mode control logic is an important element in the coder design. The parameters of the algorithm described in section 3.3 were determined experimentally. Using the peak level detector output of the signal level detector, a mode selection rate of 15 comparisons per second was found to be adequate. An active speaker threshold of 10 and an inactive speaker threshold of 0 gave good performance. Somewhat more rapid high-to-low quality switching was found desirable than low-to-high quality. An increment of 1 and a decrement of 2 resulted in approximately a 0.4 second delay for the high-to-low quality transition when the primary speaker was interrupted and approximately a 0.8 second low-to-high quality transition when the primary speaker regained the floor. Precise determination of the optimal switching delays requires evaluation in a conversational environment in which the speech output of all participants is coded.

Available time did not permit a detailed evaluation of network imposed delays. However, the preferred switching delays are relatively long. Therefore, controlling a widely distributed conference where the transmission of control information is delayed due to the presence of satellite links is not expected to cause significant problems.

The main conclusion of this study is that dynamic channel assignment based on voice activity is technically feasible for use in the NCATS environment. A dual-mode coder based on predictive coding techniques has shown itself to be particularly attractive for this application.

REFERENCES

1. J.D. Markel and A.H. Gray, "Linear Production of Speech", Springer-Verlag, 1976.
2. J.M. Turner, "Adaptive Predictive Coding for Speech Transmission at 16 kbps", TM 32057, Bell-Northern Research, Nov. 1979.
3. B.S. Atal and S.L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", J. Acoust. Soc. Am., vol. 50, pp. 637-655, Aug. 1971.
4. J.D. Markel, "The SIFT Algorithm for Fundamental Frequency Estimation", IEEE Trans. Audio Electroacoust., vol. AU-20 pp. 367-377, Dec. 1972.
5. L.J. Retallack, C.W. Reedyk and W.R. Akam, "Extension Service in Digital Telephony", International Conference on Communications, pp. 31.3.1-31.3.4, 1981.